



Market Structure, Scale Economies and Industry Performance

Amir, Rabah

Publication date:
2000

Document version
Early version, also known as pre-print

Citation for published version (APA):
Amir, R. (2000). *Market Structure, Scale Economies and Industry Performance*. Department of Economics, University of Copenhagen.

DISCUSSION PAPERS
Department of Economics
University of Copenhagen

00-08

Market Structure, Scale Economies
and Industry Performance

Rabah Amir

Stu­di­estræde 6, DK-1455 Copenhagen K., Denmark
Tel. +45 35 32 30 82 - Fax +45 35 32 30 00
<http://www.econ.ku.dk>

Market Structure, Scale Economies and Industry Performance*

Rabah Amir[†]

February 2001

Abstract

We provide an extensive and general investigation of the effects on industry performance (profits and social welfare) of exogenously changing the number of firms in a Cournot framework. This amounts to an in-depth exploration of the well-known trade-off between competition and production efficiency. We establish that under scale economies, welfare is maximized by a finite number of firms. Our results shed light on several theoretical issues and policy debates in industrial organization, including the relationship between the Herfindahl index and social welfare, destructive competition and natural monopoly. Our analytical approach combines simplicity with generality.

Key words and phrases: Cournot oligopoly, returns to scale, entry, equilibrium comparative statics.

JEL codes: D43, D60, L13, L40.

* The author is grateful to Val Lambson for several helpful conversations on the topic of this paper. This work has also benefitted from comments from Stephen Martin, William Novshek, John Roberts, Jean Tirole, as well as feedback from seminar audiences at CORE, WZB-Berlin, and the Universities of Aarhus, Arizona, Bergen, U.L.Brussels, Cergy-Pontoise, Copenhagen, Helsinki, Manchester, Odense, Strasbourg and Stony-Brook.

[†] School of Economic Studies, University of Manchester, Manchester M13 9PL, United Kingdom. E-mail: Rabah.Amir@man.ac.uk

1. Introduction

There has been a rich and insightful debate in industrial organization about the welfare and profitability effects of increasing the number of firms in an industry. While many complex facets of this fundamental issue remain partly unsettled, the basic compromise at work is well-known, though still a source of major controversy both among academics and antitrust practitioners¹. On the one hand, conventional intuition – sometimes wrongly – holds that increasing the number of firms reduces monopoly power and allows closer approximation of the competitive ideal. On the other hand, increasing the number of firms may result in reduced ability to take advantage of scale economies.

The relationship between concentration and profit rates has been one of the most active research areas in empirical industrial organization (e.g. Mueller, 1986). While conventional intuition holds that per-firm profit must decline with the number of firms, interaction between theoretical and empirical work has over the years uncovered the potential role of several inter-related factors, including potential collusion, ease of entry, and merger policy. The importance of scale economies has been stressed early on to argue against antitrust-mandated break-ups of large firms and for a *laissez-faire* policy (e.g. Demsetz, 1973).

The welfare implications of market structure have been prominent in the early beginning of the field, Bain (1956). The unquestioned view then was that barriers to entry are responsible for the presence of imperfect competition, which in turn results in sizable welfare losses. The belief that public policy must correct for this imperfection by removing barriers to entry and possibly subsidizing entry had dominated the profession and persisted until quite recently. Perceptive work by von Weiszacker (1980), Perry (1984), Mankiw and Whinston (1986) demonstrated that this view was fundamentally ill-founded by showing that if firms' conduct is not subject to regulation, then free entry results in an excessive (endogenous) number of firms relative to a social optimum. In addition to developing a version of this result, Suzumura and Kiyono (1987) also show the same conclusion holds with respect to first-best entry regulation (also see von Weiszacker, 1980).

Once the existence of the trade-off between the benefits of competition and the cost-

¹ The current Microsoft case is a timely reminder of the importance of this issue, albeit with a relatively new key dimension: Network effects.

savings of production scale is established, the main issue becomes largely empirical. Indeed, the empirical literature on the subject is extensive, though somewhat outdated by now. Several studies in Goldschmid, Mann and Weston (1974) provide empirical evidence and a general debate on this controversial issue. For instance, Sherer (1974) concludes on the basis of a detailed empirical study of twelve manufacturing industries that the evidence provides little support for the conjecture that concentration is the result of a realization of scale economies. On the other hand, his findings are controversial (e.g. Brozen, 1973).

The present paper offers a thorough theoretical investigation of the effects on industry profits and social welfare of exogenously increasing the number of firms in a Cournot industry composed of identical firms. Given the importance that these issues have for a free market society, the results presented here would be a pre-requisite to any modern empirical work or formulation of public policy dealing with market structure and social welfare.

An extensive older literature addresses the related questions of the effects of the number of firms on industry price and output levels, e.g. Ruffin (1973) and Seade (1980)². Yet, to the best of our knowledge, no systematic theoretical analysis of industry profits and social welfare has been conducted with an exogenous number of firms.

An important aspect of the paper, from a methodological viewpoint, is its reliance on the new lattice-theoretic comparative statics approach. We build directly on Amir and Lambson (2000) who use this same framework to analyse price and output effects. They derive two main results. The first one is that industry price falls (increases) with the number of firms if a firm's residual inverse demand declines slower (faster) than its marginal cost, globally.³ This is the so-called property of quasi-competitiveness (quasi-anticompetitiveness). Strong scale economies are required for demand to decline slower than marginal cost, and lead to the counterintuitive outcome on price. The second result is that per-firm profit falls with the number of firms in both cases.

A complementary methodological feature is our reliance on tight illustrative examples. These serve a dual purpose. First, they confirm that the given sufficient conditions are, in some sense, critical to the resulting conclusions. Second, they illustrate in a more accessible

² See also Frank (1973), Okuguchi (1973) and Novshek (1980), among others.

³ This sharp intuition behind many of our main conclusion is only possible with the lattice-theoretic approach. In previous work, other method-imposed assumptions, such as concave profits and decreasing marginal revenue, prevented such simple and clear-cut economic interpretations based only on critical assumptions.

manner the interaction between the various effects at work in the comparative statics at hand. In particular, we present a Main Example (Section 3) which is a blueprint for the entire paper, in that most effects of interest can be captured by varying the Example parameters. This example can serve as a pedagogical tool to convey the main ideas of the paper in a simple and intuitive framework to undergraduate students or policy practitioners.

For industry performance comparisons (Section 4), the quasi-competitive case has two clear-cut welfare results that are independent of the returns to scale: (i) welfare increases in the number of firms whenever per-firm output does, and (ii) in case of multiple equilibria, the maximal output equilibrium is the social-welfare maximizing equilibrium. The latter result vindicates the supremacy of consumer welfare over producer welfare. Otherwise, this case gives rise to two subcases, depending on the returns to scale in production. In the presence of economies of scale, industry profits are shown to globally decline with the number of firms, while social welfare is generally not monotonic. More precisely, we argue that the slightest amount of scale economies leads to welfare being decreasing at sufficiently high number of firms. Inversely, under diseconomies of scale, social welfare is globally increasing in the number of firms, while industry profits exhibit a tendency to initially increase in the number of firms (treated as a real number), starting at monopoly level. (Whether this tendency leads to duopoly, say, having higher total profit than monopoly depends on the magnitude of the returns). As an important corollary of the two monotonicity results above, under constant returns to scale, both conventional beliefs indeed hold: Industry profits fall and social welfare increases with the number of firms.

For the quasi-anticompetitive case, monopoly always leads to the highest possible industry profits, with this being the only clear result on industry profits.⁴ On the other hand, the welfare outcome is unambiguous: Social welfare always decreases in the number of firms. In this case there are strong enough scale economies to overcome all other considerations.

Our conclusions provide a precise theoretical foundation for intuitive beliefs about the need for a trade-off between the benefits of fostering increased competition and the ability of firms to exploit scale economies. The conclusions are also fully congruent with some classical results in partial equilibrium analysis. It is well-known (Ruffin, 1973) that, under

⁴ We provide an informal argument showing that, for intermediate market structures, the cost and revenue effects move in conflicting directions, the latter in a counterintuitive way..

quasi-competitiveness, Cournot equilibria converge to perfectly competitive equilibria when average cost is nondecreasing, and that this same convergence fails when average cost is nonincreasing. Our results show that welfare is monotonically increasing in the former case (thus converging to first-best welfare), but not in the latter. Also see Novshek (1980).

The results presented here may also be invoked to illuminate a number of important theoretical and public policy debates: See Section 5 where some important corollaries of our results are presented in various applications. The first of these concerns the relationship between Cournot and perfectly competitive equilibria, as described above.

The second point deals with the comparison of endogenous concentration levels prevailing under free entry and first or second-best socially optimal entry. Mankiw and Whinston (1986) and Suzumura and Kiyono (1987) show that the old belief that free entry leads to socially too few firms is invalid under general conditions. Our results provide simple but important insights into this important issue in addition to clarifying in important ways the latter authors' conclusions on the comparison between free and first best entry.

The third point addresses the extensive use made by antitrust authorities of the Herfindahl-Hirschman index of concentration under the presumption that the Index is a good inverse measure of social welfare. Recent theoretical work showed that with a constant number of firms, any output transfers across firms that leave price unchanged must cause the Index and social welfare to move in the same direction, Farrell and Shapiro (1990) and Salant and Shaffer (1999). We complement this insight by the observation that with scale economies and a varying number of firms, both the Index and welfare decrease with the number of firms, when the latter is larger than a threshold level, which may be one.

The fourth point considers the long-standing concentration/profitability debate, and relates our results to the inconclusive evidence uncovered over the years on this key issue. In particular, in cases where free entry leads to a completely indeterminate number of firms (the quasi-anticompetitive case) or to multiple equilibria with a given number of firms, any meaningful correlation between industry characteristics and profits is unlikely.

The fifth point identifies the Cournot model in the quasi-anticompetitive case as an appropriate (noncooperative) framework for modelling the concept of destructive competition that was prevalent in the old regulation literature (e.g. Sharkey, 1982). Indeed, there is an

excellent match between the theoretical predictions of the Cournot model in that case and the stylised facts commonly associated with destructive competition.

The sixth point proposes to define natural (unregulated) monopoly as an industry where the socially optimal number of firms is one, as opposed to the old definition of (regulated) natural monopoly based on the inability to improve on costs by subdividing production, a purely production-based criterion. This new definition clearly balances the market and production sides, and is more appropriate in the absence of regulation and contestability.

All these applications emphasize the role of scale economies in engendering a trade-off between the market effect and the production efficiency effect. They convey the sense that our simple results form a pre-requisite for a thorough understanding of the issues presented.

2. The Model

The fundamental questions under investigation here can be simply phrased as follows: How do total equilibrium output (and hence industry price), per-firm profit, industry profit and social welfare vary with the *exogenously given* number of firms in the industry? We consider these fundamental questions in the framework of equilibrium comparisons (as in Milgrom and Roberts, 1994), the exogenous parameter being the integer number of firms.

We begin with the basic notation. Let $P : R^+ \rightarrow R^+$ be the inverse demand function, $C : R^+ \rightarrow R^+$ the (common) cost function, $A : R^+ \rightarrow R^+$ the average cost function, and n the number of firms in the industry. Let x denote the output of the firm under consideration, y the total output for the other $(n - 1)$ firms, and z the cumulative industry output, i.e., $z = x + y$. At equilibrium, these quantities will be indexed by the underlying number of firms n . We explicitly deal with the (possible) nonuniqueness of Cournot equilibria by considering extremal equilibria. Denote the maximal and minimal points of any equilibrium set by an upper and a lower bar, respectively. Thus, for instance, \bar{z}_n and \underline{z}_n are the highest and lowest total equilibrium outputs, with corresponding equilibrium prices \underline{p}_n and \bar{p}_n , respectively. Performing comparative statics on equilibrium sets will consist of predicting the direction of change of these extremal elements as the exogenous parameter varies.

The profit function of the firm under consideration is

$$\Pi(x, y) = xP(x + y) - C(x) \tag{1}$$

Alternatively, one may think of the firm as choosing total output $z = x + y$, given the other firms' cumulative output y , in which case its profit is given by

$$\tilde{\Pi}(z, y) = \Pi(z - y, y) = (z - y)P(z) - C(z - y) \quad (2)$$

Let $\Delta(z, y)$ denote the cross-partial derivative of $\tilde{\Pi}$ with respect to z and y , i.e.,

$$\Delta(z, y) = -P'(z) + C''(z - y) \quad (3)$$

Note that both $\tilde{\Pi}$ and Δ are defined on the set $\varphi \triangleq \{(z, y) : y \geq 0, z \geq y\}$.

The following **Standard Assumptions** are in effect throughout the paper:

(A1) $P(\cdot)$ is continuously differentiable and $P'(\cdot) < 0$.

(A2) $C(\cdot)$ is twice continuously differentiable on $(0, \infty)$ and $C'(\cdot) > 0$.

(A3) There exists $\hat{x} > 0$ such that $P(x) < A(x)$ for all $x > \hat{x}$.

Although convenient, the smoothness assumptions are by no means necessary for our main results. (A3) simply guarantees that a firm's reaction curve eventually coincides with the horizontal axis, so that a firm's effective outputs, and thus all Cournot equilibrium outputs, are bounded by some constant, say K , for all n .

The qualitative nature of most of our results hinges entirely on the global sign of Δ , so that we will distinguish two main cases: $\Delta > 0$ and $\Delta < 0$. When $\Delta > 0$ globally, there will be two subcases of interest depending on the returns to scale, or the slope of the average cost curve. This division is already apparent in the upcoming example, which may serve as a blueprint for the entire paper.

3. Concentration, Returns to Scale and Industry Performance: The Main Example

We now consider a simple example that provides an excellent and thorough overview of most of the results derived in this paper.⁵ As a parameter capturing the returns to scale is varied, the example can fit the two major cases of analysis of the general model: $\Delta < 0$ and $\Delta > 0$. In the latter case, the example can also capture the two subcases of interest: economies or

⁵ This example would be very appropriate for the purpose of presenting in a very elementary framework the essentials of the analysis to undergraduate students or economic practitioners.

diseconomies of scope. In addition, this example will also be invoked later on to gain further insight into the tightness of the conditions behind our general results.

Let the inverse demand be linear and the cost function be quadratic⁶, i.e.,

$$P(z) = a - bz \quad \text{and} \quad C(x) = cx + dx^2 \quad (4)$$

with the assumption throughout that $a > c > 0$, $b > 0$, $b + d > 0$.

Since $A(x) = c + dx$, returns to scale are increasing (decreasing) if $d > (<)0$. Thus d is our returns to scale parameter, key to many results below. Furthermore (cf. (3))

$$\Delta = b + 2d \gtrless 0 \text{ if } d \gtrless -b/2 \quad (5)$$

The reaction function is always linear (when strictly positive) and given by $r(y) = \frac{a-c-by}{2(b+d)}$. Thus for any number of firms n there is always a unique symmetric Cournot equilibrium. Omitting some lengthy calculations (including solving for the symmetric Cournot equilibrium via $r[(n-1)x_n] = x_n$), this equilibrium satisfies

$$x_n = \frac{a-c}{b(n+1)+2d}, \quad \pi_n = \frac{(b+d)(a-c)^2}{[b(n+1)+2d]^2}, \quad W_n = \frac{n[b(n+2)+2d](a-c)^2}{2[b(n+1)+2d]^2} \quad (6)$$

Furthermore, if $d > -b/2$ (or $\Delta > 0$), the slope of the reaction curve is larger than -1 and the symmetric equilibrium is the unique equilibrium. It is also globally stable in the sense that best-reply Cournot dynamics converges to this equilibrium, from any initial outputs.

On the other hand, if $d < -b/2$ (or $\Delta < 0$), then $r'(y) \leq -1$, so that $r(y)$ decreases steeply and is equal to 0 when $y \geq (a-c)/b$.⁷ For the n -firm oligopoly, the unique symmetric equilibrium (with all firms active) is unstable in the sense that best-reply Cournot dynamics diverge away from it (Seade, 1980).

It can be verified that for the unique symmetric equilibrium:

(i) per-firm output x_n is always decreasing in n (cf. Proposition 1b).

⁶ This specification of the demand and cost functions with the restriction $d < 0$ has already been considered with different motivations in Cox and Walker (1998) and d'Aspremont, Gerard-Varet and Dos-Santos-Ferreira (2000). Neither study deals with our main issues: industry profits and social welfare.

⁷ Consequently, there are several other Cournot equilibria, and they can be characterized as follows. With n being the total number of firms in the industry, if any m firms (with $m < n$) produce the output x_m each (given in (6)), and the remaining $n-m$ firms produce nothing, the resulting output configuration is clearly a Cournot equilibrium. To see this, observe that $r(mx_m) = 0$, since $mx_m \geq (a-c)/b$, as can be easily checked. In particular, if any one firm produces the optimal monopoly output $x_1 = (a-c)/(b+2d)$, and all the others produce nothing, we have a Cournot equilibrium. Again, this follows from $r(x_1) = 0$ since $x_1 \geq (a-c)/b$, and $x_1 = r(0)$ clearly. Given the linearity of the reaction curve here, this is easy to see graphically.

(ii) industry output $z_n = nx_n$ is increasing in n if $d > -b/2$ (or $\Delta > 0$) and decreasing in n if $d < -b/2$ (or $\Delta < 0$) (cf. Propositions 1a and 2a).⁸

(iii) per-firm profit π_n is always decreasing in n (cf. Propositions 1c and 2c).

It remains to analyse the effects of n on industry profits and social welfare. It is convenient here to treat the number of firms as a real variable. For industry profits, we have (with the computational details left out)

$$\frac{\partial(n\pi_n)}{\partial n} \begin{matrix} \geq \\ \leq \end{matrix} 0 \text{ if and only if } n \begin{matrix} \leq \\ \geq \end{matrix} 1 + 2d/b. \quad (7)$$

Here, there are two separate cases of interest:

(i) $d < 0$: Then industry profits always decrease with the number of firms, with in particular monopoly having the largest industry profit.

(ii) $d > 0$: Then $\tilde{n} = 1 + 2d/b$ maximizes industry profits, which thus increase in n when $n < \tilde{n}$ and decrease in n when $n > \tilde{n}$, starting from any given n . Observe that if $d/b < 1/2$, then $1 < \tilde{n} < 2$. Hence, in particular, if $\tilde{n} = 1$ (i.e. monopoly is the market structure that maximizes total profits), industry profits would be globally decreasing in n . But if d is large enough, i.e., if there are sufficiently high returns to scale, industry profits will be rising in n initially, all the way to \tilde{n} which may be a large number of firms, but industry profits always eventually decrease in n (i.e., for $n > \tilde{n}$.)

For social welfare, one can easily verify that

$$\frac{\partial W_n}{\partial n} \begin{matrix} \geq \\ \leq \end{matrix} 0 \text{ if and only if } nbd \begin{matrix} \geq \\ \leq \end{matrix} -(b+d)(b+2d) \quad (8)$$

Again, there are two separate cases of interest:

(a) $\Delta < 0$ iff $d < -b/2$: Welfare always decreases with n (cf. Proposition 10.)

(b) $\Delta > 0$ iff $d > -b/2$: Here, there are two different subcases of interest.

(i) $d > 0$: welfare always increases in n (cf. Proposition 6.)

(ii) if $-b/2 < d < 0$, then welfare is maximized at $n^* = -(b+d)(b+2d)/bd$, increases in n for $n < n^*$ and decreases in n for $n > n^*$ (cf. Proposition 8). Observe that this statement is true no matter how close d is to 0 (from below)! In other words, the slightest presence of

⁸ The intuition behind the counter-intuitive case $\Delta < 0$ is that with more competition, each firm lowers output drastically since $r'(y) < -1$, thereby moving up its steeply declining average cost curve. The resulting efficiency loss is large enough to overcome the downward pressure on price engendered by the increase in competition. The increase in average cost is passed on to consumers via a higher price.

(uniform) scale economies causes welfare to be eventually declining in n (i.e., for sufficiently large values of n). The parameters of this example can be chosen to make n^* equal any desired value from 1 on, while satisfying all the underlying constraints here.

The economic intuition can now be stated concisely and precisely since the main results hinge mainly on the sign of $\Delta = -P' + C''$, and sometimes also on the returns to scale. For industry price, there are two effects at work, a market or competition effect captured by the term $-P'$, and a production or scale effect captured by C'' . The market effect always pushes in the intuitive direction that price should fall with the number of firms. The scale effect goes in the same direction if and only if costs are convex. When costs are concave, the overall outcome on price is determined by the relative strength of the two effects. Per-firm profit always behave in the intuitive way.

For industry profits, the market effect pushes in the intuitive direction if and only if industry price is well-behaved ($\Delta > 0$). The production effect works in the intuitive direction if scale economies are present. When the two effects are antagonistic, the outcome depends on the relative strengths again.

Viewed as the sum of consumer and producer surpluses, social welfare can be discussed on the basis of the previous assessments. Thus, with $\Delta > 0$ and diseconomies of scale, consumer surplus increases with n , overcoming a possible decrease in producer surplus (the latter effect being ambiguous). With $\Delta < 0$, strong scale economies are necessarily present, and both surpluses decrease with n . Finally, with $\Delta > 0$ and economies of scale, consumer surplus moves up and producer surplus down, with an ambiguous net effect.

In conclusion, this example provides a microcosm for the entire paper. In the remainder, we present a generalization of the insights illustrated so far, preserving another key role for this Example in testing the tightness of the sufficient conditions given for our various results.

4. A General Cournot Analysis of Industry Performance

This section contains the general analysis of the interplay between market structure or concentration and scale economies in determining industry performance as reflected in price, outputs, industry profits and social welfare. This amounts to comparing Cournot equilibria along these characteristics as the number of firms varies. In an attempt to obtain the broadest

possible understanding of the issues involved, we provide a series of minimally sufficient conditions for the desired conclusions, combined with tight complementary examples to shed further light on the relationship between assumptions and conclusions. The proofs combine analytical simplicity with generality. Nonetheless, we supplement the presentation with some heuristic arguments whenever it is felt they may provide additional insight.

Methodologically, we make crucial use of the lattice-theoretic comparative statics approach.⁹ This allows for very general conclusions relying only on critically needed assumptions, thereby leading to clean and tight economic interpretations of the conclusions, as well as analytical rigor. In the present context, the usual arguments in favor of this approach become even more pertinent, as the parameter of interest, the number of firms, is an integer: See Appendix for a graphical illustration of many important details on this issue.¹⁰

While the condition $\Delta > 0$ is familiar in Cournot theory at least since Hahn (1962), it has typically been used in conjunction with many other assumptions, such as some form of concavity of each firm's profit in own output, decreasing marginal revenue, etc...The latter assumptions interfere with a good intuitive understanding of the economic forces at work, as they are also made for the case $\Delta < 0$. As shown below, there is a very natural division of the results here, and it depends *only* on the global sign of Δ . The latter has a very simple and appealing interpretation: $\Delta > 0$ ($\Delta < 0$) means that price, or residual inverse demand, decreases (increases) faster than marginal cost. Since $P' < 0$, it is clear that the convexity of C implies $\Delta > 0$ on φ . Likewise, strong concavity of C is required for $\Delta < 0$. Examples provided below illustrate that $\Delta > 0$ can hold globally even when the cost function is everywhere concave, an important subcase of analysis in this paper.

⁹ In particular, we invoke the general results of Milgrom and Roberts (1994), in addition to Topkis (1978). More specific to Cournot oligopoly, we build on the results of Amir and Lambson (2000). See also McManus (1962) and Roberts and Sonnenschein (1976) for early antecedents.

¹⁰The traditional method based on the Implicit Function Theorem and the signing of derivatives can provide insight for special cases, but is ill-suited for the requisite analysis at hand. The main reason is that it rests on assumptions (such as concavity and equilibrium uniqueness) that need not be satisfied in our general setting. Furthermore, as these same assumptions are needed in all the otherwise mutually exclusive cases of analysis, this method prevents a tight intuitive understanding of the economic forces behind each result. Finally, in the context at hand, as the parameter of interest is an integer, traditional methods have further drawbacks: See de Meza (1985), Amir and Lambson (2000) and Milgrom and Roberts (1994).

4.1 Equilibrium Price and Outputs

The results of this subsection have been proved in Amir and Lambson (2000). They are stated here without proof, interpreted and then used in the sequel in looking at industry profit and social welfare. For a more detailed presentation, the reader is referred to the above paper. In the Appendix, a graphical illustration of the need for the new comparative statics is presented, with the conclusion that only extremal equilibria can be unambiguously compared as the number of firms varies.¹¹ For any variable of interest, the maximal (minimal) value will always be denoted by an upper (lower) bar.

Proposition 1 *Let $\Delta(z, y) > 0$ on φ . For each n , there exists a symmetric equilibrium and no asymmetric equilibria. Let x_n be an extremal Cournot equilibrium output.*

- (a) *Industry output z_n is nondecreasing in n , and hence price p_n is nonincreasing in n .*
- (b) *x_n is nonincreasing [nondecreasing] in n if $\log P$ is concave [convex and $C(\cdot) \equiv 0$].*
- (c) *The corresponding equilibrium profit π_n is nonincreasing in n .*

Thus the Cournot model is quasi-competitive here (Part (a)). The fundamentally needed assumption is the supermodularity of $\tilde{\Pi}$ on φ , which is equivalent to $\partial^2 \tilde{\Pi} / \partial z \partial y = \Delta > 0$. This implies that the line segment joining any two points on the graph of the reaction correspondence r of a firm must have a slope ≥ -1 , which means that, in response to an increase in rivals' output, a firm can never contract its output by more than this increase. In particular, this precludes downward jumps for r (while allowing for upward jumps).¹²

The other case is characterized by the assumption $\Delta(z, y) < 0$, implying that r has all its slopes bounded above by -1 : as the joint output of the rivals is increased, a firm optimally reacts by contracting its output so much that the resulting total output decreases.

Proposition 2 *Let $\Delta(z, y) < 0$ on φ . Then assuming $\Pi(\cdot, y)$ is quasi-concave, we have:*

- (a) *There is a unique symmetric equilibrium, and it satisfies: x_n , z_n and π_n are nonincreasing in n . Hence p_n is nondecreasing in n .*
- (b) *For any m with $1 \leq m < n$, the following is an equilibrium for the n -firm oligopoly: Each of any m firms produces x_m while the remaining $(n - m)$ firms produce nothing. All these Cournot equilibria are invariant in n , in that all entering firms would produce zero.*
- (c) *There are no other Cournot equilibrium (than those of Parts (a) and (b)).*

For this case, we henceforth focus on the symmetric equilibria¹³ (from Part (a)).

¹¹For unstable equilibria (in the sense of best-reply Cournot dynamics), the price comparative statics is counter-intuitive (as seen in the Appendix), and this will carry through to other results.

¹²This property was noticed and exploited in classic papers by McManus (1962) and Roberts and Sonnenschein (1976) to establish the existence of Cournot equilibrium in the case of symmetric firms with convex cost functions. See Amir and Lambson (2000) for a generalization.

¹³

4.2 Industry Profits and Social Welfare

Here, the effects of an exogenous change in the number of firms on total profits and social welfare are investigated. Most of the results below are stated both in a local and in a global sense, the latter being a direct consequence of the former. Elementary (heuristic) proofs for the global statement are given in the text, while rigorous proofs for the local statement are given in Appendix. We emphasize that the latter, while not standard, nonetheless combine generality with *simplicity!* For the same reasons as before, we continue to focus on the two extremal equilibria for all our results, and to separate the analysis of our comparative-equilibria results into two cases, according to the global sign of Δ .

4.3 The Case $\Delta > 0$

Recall that $\Delta > 0$ globally is consistent with both globally increasing and decreasing returns to scale. We begin with the effects of concentration on industry profits. The local statement (a) below is more general than the global statement (b). The former requires a rigorous proof, given in Appendix (A1), while the latter is heuristically proved here with a standard approach. This separation of proofs applies to most of our main results below.

Proposition 3 *Let $\Delta > 0$ on φ . For the extremal equilibria,*

(a) *Industry profit $n\pi_n$ is globally nonincreasing in n if $A(\cdot)$ is nonincreasing.*

(b) *$n\pi_n \geq (n+1)\pi_{n+1}$ for any given n if $A(\frac{n+1}{n}x_{n+1}) \leq A(x_{n+1})$.*

Proof. We present here a simple heuristic proof of (a). Assuming n is a real and $n\pi_n = z_n[P(z_n) - A(x_n)]$ is differentiable in n , one can get $\frac{d(n\pi_n)}{dn} = \frac{dz_n}{dn}[P(z_n) - A(x_n)] + z_n[P'(z_n)\frac{dz_n}{dn} - A'(x_n)\frac{dx_n}{dn}]$. Using the first-order condition for a symmetric Cournot equilibrium,

$$P(z_n) - A(x_n) + x_n[P'(z_n) - A'(x_n)] = 0 \quad (9)$$

and simplifying, one arrives at

$$\frac{d(n\pi_n)}{dn} = x_n A'(x_n) + (n-1)P'(z_n)\frac{dz_n}{dn} \quad (10)$$

Now, $\frac{d(n\pi_n)}{dn} \leq 0$ follows from $A'(x_n) \leq 0$, and $\frac{dz_n}{dn} \geq 0$ since $\Delta > 0$ (Proposition 1a). ■

Here existence of a symmetric equilibrium is not guaranteed for all n without the quasi-concavity assumption, as the best-response may have a downward jump where it skips over the $y/(n-1)$ lines for some n 's, thus implying the absence of a symmetric equilibrium for those n 's.

Since per-firm profit π_n always falls with n , Proposition 3 asks whether π_n falls fast enough to have $n\pi_n \geq (n+1)\pi_{n+1}$. In interpreting the proposition, it is convenient to separate the overall effect of an increase in the number of firms on industry profits into two distinct parts, as suggested by the above proof. The market or total revenue effect, which may be isolated by setting $A' = 0$, always pushes in the intuitive direction that industry profits must fall. On the other hand, the production or efficiency effect goes in the same direction if and only if scale economies are present.¹⁴

The proof of the result also makes it clear that the conclusion follows when both the market and the production efficiency effects push in the same direction, which suggests the condition on average cost is sufficient but not necessary. It is natural then to ask how critical this assumption is for this conclusion. Treating the number of firms as a real number, the following argument provides a simple but interesting insight: Monopoly is never the profit maximizing market structure under increasing average cost.

Proposition 4 *if $A'(\cdot) > 0$, industry profit increases from monopoly level as the number of firms is increased slightly beyond $n = 1$.*

Proof. Setting $n = 1$ in (10), we have $\left[\frac{d(n\pi_n)}{dn} \right]_{n=1} = x_1 A'(x_1) > 0$. ■

Observe that this need not mean that duopoly has higher profit than monopoly, as industry profit may peak between $n = 1$ and $n = 2$, with either π_1 or $2\pi_2$ as the highest value. This point is illustrated in the Main Example where, for $d < b/2$, industry profits may well be globally decreasing in the number of firms, and are certainly decreasing in n for $n \geq 2$ (see (7)). Thus, Proposition 4 relies crucially on the number of firms being a real.

Nonetheless, the point made here is important as it shows that the slightest amount of increasing returns pushes toward industry profits that are increasing in the number of firms. Whether this effect actually succeeds in preventing industry profit from being globally decreasing in the *integer* number of firms depends on the strength of the increasing returns, as suggested by the Main Example. Indeed, from (7), a sufficient condition for industry profit not to be globally decreasing in n is $d > b/2$.

¹⁴By contrast, an n -firm cartel always has higher optimal profit than the total n -firm oligopoly profit, since the cartel, with access to n plants, always has the option of producing nx_n at a cost at most equal to the total cost of the n -firm oligopoly. There is thus an obvious difference between a monopoly (with access to one plant) and a cartel composed of n identical firms.

We now turn to the welfare analysis. In case of multiple Cournot equilibria, \underline{x}_n is the Pareto-dominant equilibrium for the firms (i.e. leads to the largest producer surplus) and Pareto-worst equilibrium for consumers (i.e. leads to the smallest consumer surplus), while \bar{x}_n is Pareto-preferred for consumers and Pareto-worst for the firms. It is then of interest to know whether the Cournot equilibria are ranked according to (the Marshallian measure of) social welfare, defined as $\int_0^z P(t)dt - nC(z/n)$. In other words, is one of the two surpluses always dominant? The next result (proof in Appendix A1) settles this question in favor of consumer surplus.

Proposition 5 *Let $\Delta > 0$, and x_n and x'_n denote two distinct equilibrium per-firm outputs with corresponding social welfare levels W_n and W'_n . If $x_n \leq x'_n$, then $W_n \geq W'_n$. Hence, \bar{x}_n is the social welfare maximizer among all equilibrium per-firm outputs.*

As Proposition 1 shows, the case $\Delta > 0$ is consistent with both x_n increasing and x_n decreasing. The implications of these two possibilities on social welfare are quite different, as reflected in the next result. Also, if the demand function does not satisfy either condition (log-concavity or log-convexity) from Proposition 1(b), then x_n will generally not be monotonic in n . Thus, in the following result, the local statement (b) of the welfare result is more general than the global statement (a).

Proposition 6 *Let $\Delta > 0$ on φ . For any n , at an extremal equilibrium,*

- (a) *Social welfare is nondecreasing in n if (i) $A(\cdot)$ is nondecreasing and $x_n \geq x_{n+1}$, or (ii) $x_n \leq x_{n+1}$.*
- (b) *$W_{n+1} \geq W_n$ for a given n if either one of the following holds: (i) $A(x_{n+1}) \leq A(x_n)$, or (ii) $x_n \leq x_{n+1}$.*

Proof. We prove Part (a) in a simple heuristic way, and leave the proof of Part (b) to Appendix. Assuming differentiability of W_n and x_n with respect to n , and differentiating through $W_n = \int_0^{z_n} P(t)dt - z_n A(x_n)$, one gets $\frac{dW_n}{dn} = P(z_n)\frac{dz_n}{dn} - \frac{dz_n}{dn}A(x_n) - z_n A'(x_n)\frac{dx_n}{dn}$, or

$$\frac{dW_n}{dn} = \pi_n + n[P(z_n) - C'(x_n)]\frac{dx_n}{dn} \quad (11)$$

Using the first-order condition for a Cournot equilibrium (9), and simplifying yields

$$\frac{dW_n}{dn} = x_n[A'(x_n)x_n - \frac{dz_n}{dn}P'(z_n)] \quad (12)$$

For (a), the conclusion follows from (11) since $P(z_n) \geq C'(x_n)$ and $\frac{dx_n}{dn} \geq 0$. For (b), the conclusion follows from (12) since $A'(x_n) \geq 0$ and $\frac{dz_n}{dn} \geq 0$ (from $\Delta > 0$). ■

Since price falls with the number of firms here, consumer surplus always increases. However, producer surplus may a priori move either way. So the proposition identifies two sufficient conditions (diseconomies of scale and decreasing per-firm output, or increasing per-firm output) implying that total profit will never decrease enough to overcome the increase in consumer welfare and result in lower social welfare. An alternative way to think of this result is as follows. Due to the increase in industry output, the sum of consumer surplus and industry revenue (i.e. total benefit or the total area under the inverse demand up to the equilibrium output) always increases with the number of firms. On the other hand, industry costs may go either way. In this perspective, Proposition 6 identifies conditions ensuring that industry costs will never increase enough to cause social welfare to overall decrease, in spite of the increase in total benefit.

Propositions 3 and 6, taken together, imply that conventional wisdom fully prevails for the case of constant returns to scale, which is widely invoked in industrial organization.

Corollary 7 *With linear cost, $C(x) = cx$, industry profit $n\pi_n$ is nonincreasing in n and social welfare W_n is nondecreasing in n , for all n , at any extremal equilibrium.*

Proof. This follows directly from Propositions 3 and 6, as average cost is constant. ■

In the presence of scale economies, it is of interest to shed further light on the welfare-maximizing number of firms n^* . Treating the number of firms as a real variable, we can provide a characterization and an interesting interpretation of n^* (cf. (8)).

Proposition 8 *Assume $\Delta > 0$ and global scale economies prevail (i.e. $A'(\cdot) \leq -\varepsilon < 0$). Then the welfare-maximizing number of firms n^* is finite and satisfies*

$$\frac{d[P(z_{n^*})]}{dn} = C'(x_{n^*}) - A(x_{n^*}) \quad (13)$$

Proof. It is convenient here to rewrite (12) as

$$\frac{dW_n}{dn} = x_n^2 \left[A'(x_n) - \frac{1}{x_n} \frac{dz_n}{dn} P'(z_n) \right]. \quad (14)$$

Differentiating the Cournot equilibrium first-order condition $P(z_n) + x_n P'(z_n) = C'(x_n)$ with respect to n and simplifying, one can evaluate

$$\frac{1}{x_n} \frac{dz_n}{dn} = \frac{P'(z_n) - C''(x_n)}{(n+1)P'(z_n) + z_n P''(z_n) - C''(x_n)} \longrightarrow 0 \text{ as } n \rightarrow \infty.$$

Hence, given the ε such that $A'(\cdot) \leq -\varepsilon < 0$, there is some N large enough such that $n \geq N \implies -\frac{1}{x_n} \frac{dz_n}{dn} P'(z_n) < \varepsilon \leq -A'(x_n)$, and $x_n > 0$. Then from (14), it follows that $\frac{dW_n}{dn} < 0$ for all $n \geq N$. This implies W_n has a maximum for $n \leq N$. Setting $\frac{dW_n}{dn} = 0$ in (14) yields $P'(z_n) \frac{dz_n}{dn} = x_n A'(x_n)$. (13) then follows from the obvious relations $C'(x_n) - A(x_n) = x_n A'(x_n)$ and $\frac{d[P(z_n)]}{dn} = P'(z_n) \frac{dz_n}{dn}$. ■

Thus, the socially optimal number of firms has a simple interpretation: The last "marginal firm" taken in lowers industry price by exactly the difference between average and marginal cost at the equilibrium per-firm output. This equalizes marginal social benefit (the sum of consumer surplus and firms' revenues) with marginal social cost (the production costs). While the fact that the socially optimal number of firms is typically finite in the presence of fixed costs is well-known, Proposition 8 is nonetheless somewhat surprising as it relies only on the slightest level of scale economies and not necessarily on any fixed costs.¹⁵

The next example shows that if $A' \leq 0$ but $A'(0) = 0$, W_n may be globally increasing in n , so that $n^* = \infty$. Hence, the assumption $A'(0) < 0$ in Proposition 8 is crucially needed:¹⁶

Example 2. Let $P(z) = 2 - z$ and $C(x) = x - .1x^3/3$, for $x \leq \sqrt{30}$.

Thus $A(x) = 1 - .1x^2/3$ and $A'(0) = 0$. There is a unique Cournot equilibrium with $x_n = 5[n + 1 - \sqrt{(n + 1)^2 - .4}]$. It can be numerically verified that W_n is increasing in n .¹⁷

4.4 The case $\Delta < 0$

Strong economies of scale are necessary for Δ to be globally negative. One feature that is known to give rise to economies of scale is the presence of (avoidable) fixed-costs. Without these, one needs a strongly concave cost function for $\Delta < 0$ to be possible.

In view of the multiplicity of Cournot equilibria described in Proposition 2 (b), free entry would give rise to every possible integer number of firms being active, with all firms producing equal outputs.¹⁸ Furthermore, the equilibrium with m active firms is also a

¹⁵With diseconomies of scale (or $A'(\cdot) > 0$), (13) holds with a \leq sign instead of the $=$ sign. In the limit as $n \rightarrow \infty$, both sides of (13) are zero, as should be the case for perfect competition (recall that marginal and average costs intersect at the latter's minimum): See Ruffin (1971).

¹⁶The Main Example also shows that any (uniform) level of scale economies, i.e. the smallest (in absolute value) $d < 0$, the conclusion that social welfare *globally* increases with the number of firms would fail as shown by (8): See Point (b)(ii) just below (8).

¹⁷ $W_n = 5x(x + 1 - \sqrt{(x + 1)^2 - .4}) - 12.5x^2(x + 1 - \sqrt{(x + 1)^2 - .4})^2 + 4.1667x(x + 1 - \sqrt{(x + 1)^2 - .4})^3$.

¹⁸This refers to the subgame-perfect equilibria of a two-game of entry where (infinitely many) firms simultaneously decide whether to enter or not at no cost in the first stage, and the entrants then engage in Cournot

Stackelberg equilibrium of a two-stage game where the m firms act as first-movers and the rest of the firms as second movers, $m \leq n - 1$: Amir and Lambson (2000), Robson (1990).

The only general result on industry profit we can offer here vindicates the conventional wisdom only about monopoly.

Proposition 9 *Let $\Delta < 0$ on φ . Industry profit is highest under a monopoly than under any other market structure, i.e., $\pi_1 \geq n\pi_n$, for all n .*

Proof. Since the cost function is concave (hence subadditive), a single firm has the *option* of producing the n -firm total Cournot output $z_n = nx_n$ at a cost lower than that of the n -firm oligopoly (i.e., $C(nx_n) \leq nC(x_n)$) for any n . The conclusion then follows. ■

While no counterexample could be found to establish that $n\pi_n$ is not always decreasing in n , the following argument suggests the conjecture might be false. Total cost is easily seen to increase in n here, but the revenue part "moves in the counterintuitive direction".¹⁹

The welfare comparative statics is unambiguous here, due to the strong scale economies: With more firms, output per firm is strongly reduced, resulting in a drastic increase in average cost. This efficiency loss overcomes any other countervailing considerations.

Proposition 10 *Let $\Delta < 0$ on φ . Then at the unique symmetric equilibrium, social welfare W_n is nonincreasing in n , for all n .*

Proof. The conclusion follows from (12) since $A'(x_n) \leq 0$ and $\frac{dz_n}{dn} \leq 0$ (from $\Delta < 0$). ■

Since price increases with the number of firms here (Proposition 2a), consumer surplus decreases. Also, as average cost and equilibrium per-firm output both decline rapidly, equilibrium total production costs increase rapidly with the number of firms here. Hence, even if total profits go up, the increase will never be sufficient (recall also that per-firm profit goes down) to overcome the fall in consumer surplus.

Corollary 11 *Let $\Delta < 0$ on φ . Then monopoly leads both to the highest producer surplus and to the highest consumer surplus levels.*

competition in the second stage upon observing the number of entrants. Lopez-Cunat (1999) analyses the differences between this entry process and the one-stage version used by Novshek (1980) among many others.¹⁹Indeed, if it were possible to have $nx_n \geq (n + 1)x_{n+1}$ while $C \equiv 0$ (which we know is impossible since $C \equiv 0$ clearly implies $\Delta > 0$), we would have $\Pi_{n+1} = x_{n+1}P[(n + 1)x_{n+1}] \geq \frac{nx_n}{n+1}P\left(\frac{nx_n}{n+1} + nx_{n+1}\right) \geq \frac{nx_n}{n+1}P\left(\frac{nx_n}{n+1} + n\frac{nx_n}{n+1}\right) = \frac{nx_n}{n+1}P(nx_n) = \frac{n\Pi_n}{n+1}$, where the first inequality is from the Cournot equilibrium property and the second from the facts that P is decreasing and $nx_n \geq (n + 1)x_{n+1}$. It would then follow that $(n + 1)\Pi_{n+1} \geq n\Pi_n$: Industry profit would be increasing in the number of firms!

Proof. The two statements follow respectively from Propositions 9 and 2a. ■

In view of the counterintuitive nature of many of the results in the case $\Delta < 0$, it is natural to ask whether these results could have any predictive value in describing imperfect competition in some real-world markets.²⁰ Experimental evidence suggests that unique (stable) Cournot equilibria are good predictors of actual behavior (Holt, 1986). By contrast, Cox and Walker (1998) report that in a symmetric Cournot game with three equilibria, a symmetric unstable one and two boundary or monopoly equilibria (cf. Proposition 2), laboratory behavior reflected no regular patterns of play that would support any of the three equilibria. Rather, play seemed to proceed along irregular cycles around the three equilibria, meaning that the players continuously exhibited large swings in their output levels, conveying a clear sense of unstable behavior. On the other hand, none of the Nash equilibrium refinements for one-shot games (such as perfection, properness, strategic stability, etc..., see e.g. Fudenberg and Tirole, 1991) can discard Cournot-unstable equilibria, although some evolutionary learning processes can. Finally, regardless of stability properties, symmetric Cournot equilibria can always be regarded as being focal (Schelling, 1960.)

4.5 Hybrid Cases

In view of the level of generality of our conclusions, the fact that the entire analysis rests essentially on one easily checked condition on the global sign of Δ is a remarkable feature. On the other hand, there are many demand-cost combinations of interest for which Δ changes signs on its domain: Hybrid cases. For these, Cournot equilibria will generally not behave in the globally monotonic ways we uncovered here. The issue of existence also needs separate attention then. De Meza (1985) provides an interesting hybrid counterexample highlighting the differences between local and global comparative statics and showing that treating n as a real number can lead to misleading results. Nonetheless, in spite of the nonmonotonic behavior of some of the variables of interest, some of the insights we developed can still be useful here. We offer an example.

Example 3. Let $P(z) = \frac{1}{z+1}$ and $C(x) = \frac{1}{2} \log(x+1)$.

It is easily checked that $\Delta(z, y)$ changes signs on φ (so that our results do not apply here),

²⁰Further discussion of this issue is provided in Section 5 where the characteristics of the Cournot equilibria here are identified with the concept of destructive competition, among other applications.

and that $\Pi(x, y) = \frac{x}{x+y+1} - \frac{1}{2} \text{Log}(x+1)$ is quasi-concave in x , for fixed y . The reaction curve is $r(y) = \sqrt{1-y^2}$, and the (unique) Cournot equilibrium is given by $x_n = \frac{1}{\sqrt{n^2-2n+2}}$. Simple calculations show that while x_n and $n\pi_n$ are decreasing in n , for all n , z_n and thus p_n are not monotonic in n (details are left out).

Social welfare is given by $W_n = \log\left(\frac{n}{\sqrt{n^2-2n+2}} + 1\right) - \frac{n}{2} \log\left(\frac{1}{\sqrt{n^2-2n+2}} + 1\right)$. In particular, $W_1 = \frac{1}{2} \ln 2 \simeq .34657$, and $W_2 = \ln(\sqrt{2} + 1) - \ln\left(\frac{1}{2}\sqrt{2} + 1\right) \simeq .34657$, so that $W_2 \simeq W_1$. Thus²¹, $n^* = \{1, 2\}$: A social planner is indifferent (at least up to 5 decimals) between monopoly and duopoly as the optimal choice!

5. On some Theoretical and Policy Implications

The results presented here lie at the heart of the modern theory of industrial organization and can, to some extent, illuminate a number of past as well as present theoretical issues and public policy debates. In particular, we relate our findings to the relationship between Cournot outcomes and perfect competition, the regulation of entry, the welfare content of the Herfindahl index, natural monopoly and destructive competition. Surprisingly, the latter two notions have not really been linked with Cournot theory in the past. We attempt to fill this gap below. In some cases, we also present some new results here.

5.1 Relationship to Perfect Competition

Ruffin (1971) showed that if the number of firms is increased with fixed demand²², Cournot equilibria converge to the perfectly competitive equilibrium under global diseconomies of scale, but not under global economies of scale. Our conclusions shed some light on this result by indicating that (i) in the former case, equilibrium welfare converges monotonically to first-best welfare, and (ii) in the latter case, although industry profits and per-firm output both monotonically converge to zero, welfare does not increase to first-best welfare, due to firms producing at increasing (and in the limit, maximal) average cost. Here, first-best welfare would involve one firm producing the entire output and pricing at marginal cost.²³

²¹In fact, viewing n as a real variable, it can be shown that W_n is single-peaked in n and achieves its maximum at $n \simeq 1.36$.

²²See Novshek (1980) for the other approach, where demand is replicated.

²³The planner's objective is then $\max\{\int_0^{nx} P(t)dt - nC(x) : n \geq 1, x \geq 0\}$. The first-order conditions are $P(nx) = C'(x)$ and $xP(nx) = C(x)$. These imply that marginal and average cost are equal, as is well-known

5.2 Free versus Socially Optimal Entry

Perry (1984), Mankiw and Whinston (1986) and Suzumura and Kiyono (1987) put to rest a long-standing belief that free-entry leads to insufficient competition by showing that the opposite conclusion holds under general assumptions, using a two-stage game where firms decide upon entry in the first period, and then compete a la Cournot in the second period. The comparison benchmark is second-best social optimum in that the regulator controls firms' entry decisions but not their market conduct. Suzumura and Kiyono (1987) differs from the other two papers in that costless entry is assumed, and two extra results are established: (i) with convex costs, free entry is excessive relative to the first-best level, where the regulator also controls pricing or market conduct, see Footnote 24 (and von Weiszacker, 1980) and (ii) the first and second-best levels of entry are generally not comparable.

Our welfare results indicate that some of Suzumura and Kiyono's conclusions are not really instructive, and potentially misleading. Indeed, with diseconomies of scale, Cournot equilibrium welfare increases monotonically with the number of firms to the first-best level, so that both the first and the second-best socially optimal numbers of firms are *infinite* (for the former, see Footnote 24.) The free-entry number of firms is also *infinite*²⁴, so that *all three entry levels are actually equal* in a trivial way.

With (strict) economies of scale²⁵, the second-best socially optimal number of firms is finite, and equal to n^* (which is 1 if $\Delta < 0$), the first-best is always 1 (see footnote 24), while the free-entry number is infinite (see Footnote 25). Hence, free entry obviously leads to socially excessive entry under both criteria. Furthermore, the level of entry is at least as high under second-best than under first-best regulation.²⁶ Thus our results substantially clarify and extend the analysis of Suzumura and Kiyono (1987).

for a first-best solution. In particular, the first-best number of firms is then 1 if $A' < 0$ and ∞ if $A' > 0$.

²⁴The latter follows from the fact that Cournot equilibria are intersections of r with the rays $y/(n-1)$, and (i) the fact that r has all its slopes greater than -1 , if $\Delta > 0$, or (ii) the fact that r is continuous if a firm's profit is quasi-concave in own output (whether $\Delta < 0$ globally or not).

²⁵These may also prevail over the relevant range of operation regardless of the properties of the "variable" cost function in the case of costly entry (with the entry cost acting as a fixed cost). Hence, our conclusions here may also apply when entry is costly.

²⁶Both the latter statements are actually new results here, as Suzumura and Kiyono assume convex costs in the comparison between the free-entry and the first-best outcomes.

5.3 The Herfindahl Concentration Index and Welfare

The Herfindahl-Hirschman Index (or HHI) of industry concentration, defined as a (normalized) sum of the squares of the firms' market shares, is the most often used quantitative assessment of industry concentration. In particular, the value of the HHI constitutes the primary indicator for antitrust authorities of market power and of the likelihood of overt or tacit collusion in a given market. The HHI is also one of the main elements of the 1982 Merger Guidelines²⁷ in determining whether a proposed merger is to be allowed.

Underlying the extensive reliance of economic law on this measure is a fundamental belief that social welfare and the HHI are always inversely related (see e.g. Dansby and Willig, 1979). Yet, this belief has recently been challenged by theoretical studies based on the Cournot model. Farrell and Shapiro (1990) establish that, with a fixed number of (nonidentical) firms, whenever industry output is unchanged following individual firm output changes, social welfare and the HHI must change in the same direction. Salant and Shaffer (1999) provide further insight into this result in the case of constant unit costs by showing that both welfare and the HHI increase if the variance of the unit costs increases in a mean-preserving way. Also see Daughety (1990).

The present paper sheds further light on this issue by considering the effects of changing the number of firms instead. Given the symmetry assumption, the HHI with n firms here is clearly given by (a constant factor of) $1/n$. Hence, the HHI decreases if and only if the number of firms increases. On the other hand, our results indicate that in the presence of scale economies (with $A' < 0$), social welfare decreases if the number of firms exceeds some socially optimal level n^* . Thus, both the HHI and welfare decrease whenever n increases beyond n^* . In particular, in industries where $n^* = 1$, the two measures would *always* produce conflicting prescriptions as the number of firms increases.

This conclusion clearly suggests that the HHI should be augmented by some measure of economies of scale in the industry that would allow appropriate balancing between the legitimate fears of market power and the desire for production efficiency.

²⁷For some historical background on these Guidelines and an exchange of views among experts, see the Symposium in the Journal of Economic Perspectives, vol. 1, 1987.

5.4 Concentration and Profitability

One of the most extensive debates in industrial organization has revolved around the alleged positive relationship between market concentration and profits. While the majority view ended up with a belief that a weak correlation exists, the issue remains somewhat controversial. In spite of the simplicity of our framework, our results suggest new possible theoretical explanations of an elementary nature as to why this issue turned out to be so complex. Consider a two-stage game of free entry followed by Cournot competition amongst the entrants. First, the possible multiplicity of Cournot equilibria leads to different entry levels or different equilibria for the same entry level. Either way, profits per-firm will differ.

Second, our results confirm in a general way that scale economies should imply a more pronounced level of this correlation (Demsetz, 1974, Dewey, 1976 and Lambson, 1987): Since industry profits increase with concentration, per-firm profits increase at an increasing rate (i.e. $\pi_n \geq \frac{(n+1)}{n}\pi_{n+1}$). Conversely, if with diseconomies of scale, we have $\pi_n \leq \frac{(n+1)}{n}\pi_{n+1}$, the correlation between profits and concentration is more likely to be weak.²⁸

Third, with strong scale economies ($\Delta < 0$), the free-entry number of firms is fully indeterminate. In particular, with no entry cost, any number of firms is possible in a subgame-perfect equilibrium, thus leading to many possible profit levels! Postulating that the actual number of firms is determined in part by historical and other random events in such markets, no clear correlation could be expected between profits and industry characteristics.

5.5 Destructive Competition

Destructive competition was a recurrent theme in older case and empirical studies of regulated industries, particularly those in the transportation sector such as railroad and trucking (see Sharkey (1982) for a historical account). It is typically associated with a combination of industry characteristics, such as strong economies of scale (often due to large fixed costs), large productive capacity, relatively easy entry, and ill-guided government subsidies. The symptoms of destructive competition in such industries include high levels of market instability, excessive capacity and widespread price discrimination, often leading to frequent

²⁸Furthermore, three other important factors of a more dynamic nature contribute to the higher correlation in the case of economies of scale: Mergers are more likely to be sought and to be allowed, entry by new firms is more difficult (in particular if fixed costs are sizable), and collusion is thus more likely.

changes in regulatory regimes, including entry regulation.

Sharkey (1982) develops a cooperative game-theoretic approach to model destructive competition, defining industry stability by the nonemptiness of the core. The results here suggest a simple and natural alternative within the noncooperative paradigm: In the absence of any regulatory interference, destructive competition can be fruitfully modelled by Cournot competition under the assumption that Δ is globally negative. Indeed, increases in competition from any pre-existing level, including in particular monopoly, result in lower consumer welfare, per-firm profit and social welfare. Thus higher competition is unambiguously detrimental to all economic agents, with even unregulated monopoly emerging as the best among market outcomes. Furthermore, and more strikingly, some aspects of reported market instability in industries thought to have undergone phases of destructive competition may be instructively linked to the *indeterminacy in the number of active firms* and the *unstable nature of the Cournot equilibria* (in the sense of divergence of best-reply dynamics), both of which are characteristics of the case Δ globally negative (see Section 4.4.)

5.6 Natural Monopoly

Following various attempts, Baumol, Panzar and Willig (1982) provided the final definition of natural monopoly: An industry with a subadditive cost function. This is the least restrictive property of a cost function that captures the notion that any amount of final output is cheaper to produce by one firm, or, in other words, subdividing production cannot possibly save on costs. This definition completely ignores the demand side of the market, which is justified in light of two special features that were dominant in the economic scene two decades ago. The first, reflecting the prevalent public policy view of the times, is that monopolies are to be regulated anyway, so that market conduct is not really an issue, leaving production efficiency as the primary concern. The second, a theoretical belief, is that if an industry has a downward-sloping average cost curve and the market is contestable, the only stable configuration will involve a single firm pricing at average cost, resulting in zero profits.

Subsequently, a near-consensus emerged, recognizing the limited real-life validity of contestable markets²⁹, and a wave of deregulation originating in the US and the UK swept

²⁹In a book review of Baumol, Panzar and Willig (1982), Spence (1983) offered an eloquent account of the currently prevailing view on contestable markets. He concludes that the benefits of this theory lie in its

through the industrialized world. In view of the need to incorporate the demand side of the market now in a revised definition of natural monopoly, the analysis of the present paper suggests an obvious alternative: An unregulated monopoly is natural if social welfare W_n is maximized by $n = 1$. According to our results, this would require scale economies of sufficient magnitude over the relevant range, but not necessarily that Δ be globally < 0 . Recall that the Main Example shows that n^* can be equal to 1 for an industry for which $\Delta > 0$ globally. This definition is clearly more restrictive than the old one³⁰, as it incorporates the market or demand side of the industry. In other words, it strikes a socially optimal balance between the detrimental effects of concentration and the cost-saving effects of size.

More generally, a natural n^* -firm oligopoly can be analogously defined by $n^* = \arg \max_n W_n$. If $\Delta < 0$, Proposition 10 implies that n^* must necessarily be equal to 1. Hence, $n^* > 1$ is not compatible with $\Delta < 0$ globally. On the other hand, it is compatible with $\Delta > 0$ and with Δ not having a uniform sign on all its domain.

5.7 Merger Policy

Proposition 3 says that with scale economies, there are industry-wide gains to mergers. However, according to Cournot theory on mergers, in a unilateral merger, these gains are generally appropriated by nonparticipating firms, except in near-monopolization cases.³¹

An important aspect of the 1984 revisions of the Merger Guidelines is their novel consideration of cost efficiency: mergers that are likely to raise prices will be allowed if the merging firms can demonstrate by "clear and convincing evidence" that the merger will lead to significant cost savings or efficiency benefits. In practice, this clause has been exploited by candidate firms to secure approval by overstating uncertain future cost gains. Yet, ignoring the important verification and burden-of-proof issues here, our conclusions are certainly in favor of this amendment to the Guidelines for industries with known scale economies, and suggest that the extent of scale economies in the industry should be taken into account, as opposed to more firm-specific claims of technological and organizational synergies. ■

thorough analysis of cost functions, and not in its empirical relevance as a theory of market structure.

³⁰Scale economies of any magnitude imply the subadditivity of the cost function, but not vice-versa.

³¹See e.g. Farrell and Shapiro (1990) and Fauli-Oller (1997).

REFERENCES

- Amir R., and V. E. Lambson (2000), "On the Effects of Entry in Cournot Markets," *Review of Economic Studies*, to appear.
- d'Aspremont, C., R. Dos Santos Ferreira, L.-A. Gerard-Varet (2000), Contestability and the Indeterminacy of Free-Entry Equilibria, *Japanese Economic Review*, to appear.
- Bain, J. (1956), *Barriers to New Competition*, Harvard University Press, Cambridge.
- Baumol, W., J. C. Panzar and R. Willig (1982), "Contestable Markets and the Theory of Industry Structure", Harcourt Brace Jovanovich, New York.
- Brozen, Y. (1973), "Concentration and Profits: Does Concentration Matter?" in J.F. Weston and S. Ornstein, eds. *The Impact of Large Firms on the US Economy*, Lexington Books, 59-70.
- Cox, J. and M. Walker (1998), "Learning to Play Cournot Duopoly Strategies", *Journal of Economic Behavior and Organization*, 36, 141-161.
- Daughety, A. (1990), "Beneficial Concentration", *American Economic Review*, 80, 1231-37.
- de Meza, D. (1985), "A Stable Cournot-Nash Industry Need not be Quasi-Competitive", *Bulletin of Economic Research*, 37, 153, 156.
- Demsetz, H. (1973), "Industry Structure, Market Rivalry and Public Policy, *Journal of Law and Economics*, 16, 1-10.
- Demsetz, H. (1974), "Two Systems of Belief about Monopoly," in H. Goldschmid et al., eds., *Industrial Concentration: The New Learning*, Boston: Little, Brown, 164-84.
- Dewey, D. (1976), "Industrial Concentration and the Rate of Profit: Some Neglected Theory," *Journal of Law and Economics*, 19, 67-78.
- Farrell, J. and C. Shapiro (1990), "Asset Ownership and Market Structure in Oligopoly", *RAND Journal of Economics*, 21, 275-292.
- Fauli-Oller, R. (1997), "On Merger Profitability in a Cournot Setting", *Economics Letters*, 54(1), 75-79.
- Frank, C. R. (1965), "Entry in a Cournot Market", *Review of Economic Studies*, 32, 245-250.
- Goldschmid, H., H. M. Mann and J.F. Weston (1974), eds., *Industrial Concentration: The New Learning*, Boston: Little, Brown.
- Fudenberg, D. and J. Tirole (1991), *Game Theory*, MIT Press: Cambridge, MA.
- Hahn, F. (1962), The Stability of the Cournot Oligopoly Solution, *Review of Economic Studies*, 29, 329-221.

- Holt, C. (1995), "Industrial Organization: A Survey of the Results of Laboratory Experiments", in *Handbook of Experimental Economics*, A. Roth and J. Kagel eds., Princeton University Press.
- Lambson, V. E. (1987), "Is the Concentration-Profit Correlation Partly an Artifact of Lumpy Technology?" *American Economic Review*, 77, 731-733.
- Lopez-Cunat, J. M. (1999), One-Stage and Two-Stage Entry Cournot Equilibria, *Investigaciones Economicas*, 23(1), 115-28.
- MacManus, M. (1962), "Number and Size in Cournot Equilibrium", *Yorkshire Bulletin of Economic and Social Research*, 14, 14-22.
- Mankiw, N.G. and Whinston, M. D. (1986), "Free Entry and Social Inefficiency", *Rand Journal of Economics*, 17, 48-58.
- de Meza, D. (1985), "A Stable Cournot-Nash Industry Need not be Quasi-Competitive", *Bulletin of Economic Research*, 37, 153-156.
- Milgrom, P. and J. Roberts (1994), "Comparing Equilibria", *American Economic Review*, 84, 441-459.
- Mueller, D. (1986), *Profits in the Long Run*, Cambridge University Press, New-York.
- Novshek, W. (1980), "Cournot Equilibrium with free entry", *Review of Economic Studies*, 47, 473-486.
- Novshek, W. (1985), "On the Existence of Cournot Equilibrium", *Review of Economic Studies*, L II, 85-98.
- Okuguchi, K. (1973), "Quasi-Competitiveness and Cournot Oligopoly", *Review of Economic Studies*, 40, 145-148.
- Perry, M. (1984), "Scale Economies, Imperfect Competition and Public Policy", *Journal of Industrial Economics*, 32, 313-330.
- Roberts, J. and H. Sonnenschein (1976), "On the Existence of Cournot Equilibrium without Concave Profit Functions", *Journal of Economic Theory*, 13, 112-117.
- Robson, A. (1990), "Stackelberg and Marshall", *American Economic Review*, 80, 69-82.
- Ruffin, R. J. (1971), "Cournot Oligopoly and Competitive Behavior", *Review of Economic Studies*, 38, 493-502.
- Salant, S. and G. Shaffer (1999): "Unequal Treatment of Identical Agents in Cournot Equilibrium", *American Economic Review*, 89, 585-604.
- Schelling, T. (1960), *The Strategy of Conflict*, Harvard University Press, Cambridge, MA.

- Seade, J. (1980a), "On the Effects of Entry", *Econometrica*, 48, 479-489.
- Sharkey (1982), *The Theory of Natural Monopoly*, Cambridge University Press, Cambridge, U.K.
- Sherer, F.M. (1974), Economies of Scale and Industrial Concentration, in H. Goldschmid et al. eds., *Industrial Concentration: The New Learning*, Boston: Little, Brown.
- Spence, M. (1983), "Contestable Markets and the Theory of Industry Structure: A Review Article", *Journal of Economic Literature*, 21(3), 981-90.
- Suzumura, K. and K. Kiyono (1987), "Entry Barriers and Economic Welfare", *Review of Economic Studies*, L IV, 157-167.
- Topkis, D. (1978), "Minimizing a Submodular Function on a Lattice", *Operations Research*, 26, 305-321.
- von Weiszacker, C. (1980), "A Welfare Analysis of Barriers to Entry", *Rand Journal of Economics*, 11, 399-420.

6. Appendix

Here, we provide (A1) a graphical illustration of the benefits of our approach to equilibrium comparisons when $\Delta > 0$ and there are multiple equilibria, and (A2) the formal proofs.

(A1) Comparing Equilibria. The following discussion refers to Figure 1. With $\Delta > 0$, the reaction curve r can never decrease at a rate larger than -1 . With n firms, the Cournot equilibria are the intersections of the reaction curve with the line $y/(n-1)$. The number of equilibria is thus 1 for $n = 1$ and $n = 4$, 3 for $n = 2$, and 5 for $n = 3$. It is easy to see that:

1) For the extremal equilibria, industry output z_n increases with n (to see this, simply draw lines of slope -1 through these equilibria and observe the outward shifts.)

2) The motion of z_n is generally indeterminate for the middle equilibria since a may go to b, c , or d ! On the other hand, if there were only one middle equilibrium for both $n = 2$ and $n = 3$, then z_n would actually decrease for this unstable equilibrium, so that price would increase as we go from 2 to 3 firms! More generally, the comparative statics of unstable equilibria goes opposite that of the extremal equilibria. In particular, price increases with n . Furthermore, since a firm's rivals' total output, y_n , decreases with n at an unstable equilibrium, per-firm profits also increase with n at such equilibria.

3) Most of the other comparisons of interest (industry profit and welfare) depend on the previous two points, as seen in our proofs. Thus, dealing only with the extremal equilibria, one avoids a lot of indeterminacy in the comparative statics.

4) Even with the artifact of treating n as a real variable, the implicit function theorem approach cannot be justified in dealing with (say) dx_n/dn (add to this that r need not even be continuous in general). In addition, signing dx_n/dn in hybrid cases (where Δ is not uniformly signed) can be misleading (de Meza, 1985).

(A2) Proofs.

Proof of Proposition 3b. Let x_n be an extremal Cournot equilibrium, and consider

$$\begin{aligned}
\pi_n &= x_n \{P(nx_n) - A(x_n)\} \\
&\geq \frac{n+1}{n} \{x_{n+1} P[\frac{n+1}{n}x_{n+1} + (n-1)x_n] - A(\frac{n+1}{n}x_{n+1})\} \\
&\geq \frac{n+1}{n} x_{n+1} \{P[\frac{n+1}{n}x_{n+1} + (n-1)\frac{n+1}{n}x_{n+1}] - A(\frac{n+1}{n}x_{n+1})\} \\
&= \frac{n+1}{n} x_{n+1} \{P[(n+1)x_{n+1}] - A(\frac{n+1}{n}x_{n+1})\} \\
&\geq \frac{n+1}{n} x_{n+1} \{P[(n+1)x_{n+1}] - A(x_{n+1})\} \\
&= \frac{n+1}{n} \pi_{n+1}
\end{aligned}$$

where the first inequality follows from the Cournot equilibrium property, the second from the facts that P is decreasing and $nx_n \leq (n+1)x_{n+1}$ (Proposition 1a), and the third from the fact that $A(\frac{n+1}{n}x_{n+1}) \leq A(x_{n+1})$. Multiplying across by n gives the conclusion.

The global statement follows directly from the local statement. ■

Proof of Proposition 4. With z denoting industry output, industry profit is given by $\Pi_n(z) = z[P(z) - A(z/n)]$. Since for fixed z , $\Pi_n(z)$ is increasing in n if $A' > 0$, the result follows from the envelope theorem, as monopoly profit $\pi_1 = \max_z \Pi_1(z)$. ■

Proof of Proposition 5. Since $\Delta > 0$ or $P'(z) - C''(z-y) < 0$ on φ , the function $W(z) \triangleq \int_0^z P(t)dt - nC(z/n)$ is concave in z , since $W''(z) = P'(z) - \frac{1}{n}C''(z/n)$. Now, consider

$$\begin{aligned}
W'_n - W_n &= \int_0^{z'_n} P(t)dt - nC(z'_n/n) - [\int_0^{z_n} P(t)dt - nC(z_n/n)] \\
&= W(z'_n) - W(z_n) \\
&\geq W'(z'_n)(z'_n - z_n), \text{ since } W \text{ is concave.} \\
&= [P(z'_n) - C'(z'_n/n)]n(x'_n - x_n) \\
&\geq 0, \text{ since } x'_n \geq x_n.
\end{aligned}$$

The second statement of the lemma follows directly. ■

Proof of Proposition 6b. To prove Part (b)(i), consider:

$$\begin{aligned}
W_{n+1} - W_n &= \left\{ \int_0^{z_{n+1}} P(t)dt - z_{n+1}A(x_{n+1}) \right\} - \left\{ \int_0^{z_n} P(t)dt - z_nA(x_n) \right\} \\
&= \int_{z_n}^{z_{n+1}} P(t)dt - z_{n+1}A(x_{n+1}) + z_nA(x_n) \\
&\geq (z_{n+1} - z_n)P(z_{n+1}) - z_{n+1}A(x_{n+1}) + z_nA(x_n) \\
&= z_{n+1} [P(z_{n+1}) - A(x_{n+1})] - z_n [P(z_{n+1}) - A(x_n)] \\
&\geq z_{n+1} [P(z_{n+1}) - A(x_{n+1})] - z_n [P(z_{n+1}) - A(x_{n+1})] \\
&= (z_{n+1} - z_n) [P(z_{n+1}) - A(x_{n+1})] \geq 0,
\end{aligned}$$

where the first inequality follows from the fact that $P(\cdot)$ is decreasing, the second from the assumption $A(x_n) \geq A(x_{n+1})$, while the last follows from the facts that $z_{n+1} \geq z_n$ (since $\Delta > 0$) and x_{n+1} is a symmetric Cournot equilibrium.

To prove Part (b)(ii), we begin with two preliminary observations. First, the function $V_n(x) \triangleq \int_0^{nx} P(t)dt - nC(x)$ is concave in x for each n since $V_n''(x) = n[nP'(nx) - C''(x)] < 0$, as a result of $\Delta > 0$. Second, since $z_{n+1} = (n+1)x_{n+1}$ and P is decreasing,

$$\int_0^{z_{n+1}} P(t)dt = \int_0^{nx_{n+1}} P(t)dt + \int_{nx_{n+1}}^{z_{n+1}} P(t)dt \geq \int_0^{nx_{n+1}} P(t)dt + x_{n+1}P(z_{n+1}). \quad (15)$$

Now, consider,

$$\begin{aligned}
W_{n+1} - W_n &= \left\{ \int_0^{z_{n+1}} P(t)dt - (n+1)C(x_{n+1}) \right\} - \left\{ \int_0^{z_n} P(t)dt - nC(x_n) \right\} \\
&\geq x_{n+1}P(z_{n+1}) - C(x_{n+1}) + \left\{ \int_0^{nx_{n+1}} P(t)dt - nC(x_{n+1}) \right\} - \left\{ \int_0^{nx_n} P(t)dt - nC(x_n) \right\} \\
&= \pi_{n+1} + V_n(x_{n+1}) - V_n(x_n) \\
&\geq \pi_{n+1} + V_n'(x_{n+1})(x_{n+1} - x_n) \\
&= \pi_{n+1} + n[P(nx_{n+1}) - C'(x_{n+1})](x_{n+1} - x_n) \geq 0,
\end{aligned}$$

where the first inequality follows from (15), the second from the concavity of V_n in x , and the last from the facts that $P(nx_{n+1}) \geq P[(n+1)x_{n+1}] \geq C'(x_{n+1})$ and $x_{n+1} \geq x_n$. ■

Proof of Proposition 10. Consider (with $z_n \geq z_{n+1}$ here, since $\Delta < 0$):

$$\begin{aligned}
W_n - W_{n+1} &= \left\{ \int_0^{z_n} P(t)dt - z_n A(x_n) \right\} - \left\{ \int_0^{z_{n+1}} P(t)dt - z_{n+1} A(x_{n+1}) \right\} \\
&= \int_{z_{n+1}}^{z_n} P(t)dt - z_n A(x_n) + z_{n+1} A(x_{n+1}) \\
&\geq (z_n - z_{n+1})P(z_n) - z_n A(x_n) + z_{n+1} A(x_{n+1}) \\
&= z_n [P(z_n) - A(x_n)] - z_{n+1} [P(z_n) - A(x_{n+1})] \\
&\geq z_n [P(z_n) - A(x_n)] - z_{n+1} [P(z_n) - A(x_n)] \\
&= (z_n - z_{n+1}) [P(z_n) - A(x_n)] \geq 0
\end{aligned}$$

where the first inequality follows from the fact that $P(\cdot)$ is decreasing, the second from the facts that $x_n \geq x_{n+1}$ and $A(\cdot)$ is nonincreasing (the latter follows since $\Delta < 0$ requires concavity of C), and the last from the fact that $z_n \geq z_{n+1}$ (since $\Delta < 0$). ■